Queensland University of Technology
Brisbane Australia

This is the author's version of a work that was submitted/accepted for publication in the following source:

Nguyen Thanh, Kien, Fookes, Clinton B., & Sridharan, Sridha (2010) Robust mean super-resolution for less cooperative NIR iris recognition at a distance and on the move. In *SoICT '10 Proceedings of the 2010 Symposium on Information and Communication Technology*, ACM, Hanoi University of Technology, Hanoi, pp. 122-127.

This file was downloaded from: http://eprints.qut.edu.au/40916/

Notice: *Changes introduced as a result of publishing processes such as copy-editing and formatting may not be reflected in this document. For a definitive version of this work, please refer to the published source:*

Robust Mean Super-resolution for less cooperative NIR iris recognition at a distance and on the move

Kien Nguyen(1), Clinton Fookes(2), Sridha Sridharan(3)
Image and Video Research Laboratory, Queensland University of Technology
2 George Street, GPO Box 2434, Brisbane, Queensland 4001
(1) kien.nguyenthanh@student.qut.edu.au, (2),(3) {c.fookes,s.sridharan}@qut.edu.au

ABSTRACT
Less cooperative iris identification systems at a distance and on the move often suffers from poor resolution. The lack of pixel resolution significantly degrades the performance of iris recognition systems. Super-resolution has been considered to enhance resolution of iris images. This paper proposes a pixel-wise super-resolution technique to reconstruct a high resolution iris image from a video sequence of eye images. The proposed approach fuses information details from multiple frames using robust mean. Experiments on MBGC NIR portal iris video database show the validity of the proposed approach in comparison with other resolution enhancement techniques.

Keywords
Iris recognition, uncooperative human identification at a distance, super-resolution, robust mean, signal-level fusion, MBGC.

1. INTRODUCTION
Biometrics is a reliable method for the automatic identification of individuals based on their physiological and behavioral characteristics such as face, fingerprint, palmprint, gait, iris, retina, and voice. Among all the biometrics, iris has shown to be one of the most accurate traits for human identification due to its richness and stability in texture [1]. Many researchers are interested in making iris recognition less-constrained, on-the-move and at-a-distance. The most challenging problems with less cooperative iris identification at a distance and on the move are the lack of pixel resolution and noise interference such as out of focus, motion blur, frame interlacing. Super-resolution technique has been employed to address the low resolution problems.

Super-resolution is an image processing technique that reconstructs or learns lost high-frequency information to enhance the resolution of an imaging system. Super-resolution techniques can be categorized into two classes: reconstruction-based and learning-based methods. The former reconstructs lost high-frequency information by taking advantage of multiple low resolution frames of the same scene. In contrast, the latter attempts to guess the lost high-frequency information from pre-trained templates. Both methods have been utilized extensively in face image enhancement. Recently, super-resolution techniques have been considered for iris resolution enhancement.

Kwang *et al.* propose a learning-based super-resolution method based on multiple MLPs (multi-layer perceptrons) for iris recognition [2]. The proposed method restores a single low resolution image into a single high resolution image by using bilinear interpolation based on the output pixel values of the trained multiple MLPs. The middle and high frequency components of a low resolution iris image can be restored from learnt neural network architecture. Recently, Tieu Tan's group proposes another learning-based method based on CSF (Circular Symmetric Filter) [3]. Their algorithm predicts the prior relation between iris feature information of different bands and incorporates this learnt prior into the process of iris image enhancement. Both Tieu Tan's group and Kwang *et al.* methods are reported to show good performance in visual and recognition enhancement. However, the robustness of iris recognition is high degree of freedom, which means high randomness in iris images; the learning process can introduce fake high frequencies, which may mid-lead recognition procedure. In addition, both methods are conducted in artificially-created low resolution images (low resolution iris images are produced by degrading high resolution images with Gaussian kernel and down-sampling), whether it will work with real low resolution images requires more experiments.

Falmy proposes a reconstruction-based super-resolution technique to restore multiple low-resolution iris frames captured at 3 feet away [4]. The process of building a high resolution image is based on an autoregressive signature model between successive low resolution images in filling the sub pixels in the constructed high resolution image. However, the idea of using whole eye image for registration is not reasonable due to iris dilation and contractibility properties. The situation will be worst in less constrained iris recognition applications.

To overcome above dilemma when applying super-resolution to iris recognition, this paper proposes a new reconstruction-based super-resolution technique for low-resolution iris video sequences. Instead of the whole eye iris images, normalized polar-coordinate iris images will be used for super-resolution to deal with iris dilation and contraction issues. The process of building a super-resolved image from low resolution sequence is based on robust mean reconstruction technique. Robust mean reconstruction technique can incorporate middle and high frequency components from multiple low resolution frames into one desired super-resolved frame without introducing fake high frequency components. In addition, robust mean technique with a reasonable standard deviation will prevent reconstruction process from corruption by extreme pixel intensity values.

According to fusion aspect, a super-resolution technique can also be considered as a signal-level fusion technique. So here it's also worth discussing Patrick Flynn's group signal-level fusion approach by averaging multiple frames [5]. They claim to be the first to work on signal-level fusion for iris video sequences. This iris image quality enhancement technique has shown good performance on close distance MBGC iris video sequence. It will be re-implemented on long range iris database to compare with proposed approach.

Multiple Biometric Grand Challenge (MBGC), which is organized by National Institute of Standards and Technology (NIST), provides near-infrared (NIR) face portal video recorded when participants walked through a portal located 3m from a fixed-focal-length NIR camera. Iris regions can be extracted from each frame for human identification against high resolution NIR iris still template images. The database consists of 628 NIR face portal video sequences and 8589 NIR iris still images of 129 participants. This less-cooperative at-a-distance and on-the-move iris database is a challenging database since the quality of video frames is invariant with out-of-focus, motion blur, frame interlacing; iris region can be interfered severely by reflection, glasses, eyelids, eyelashes, shadows; participants close or blink their eyes in a number of frame. Examples of bad quality eye images can be found in the Figure 1.

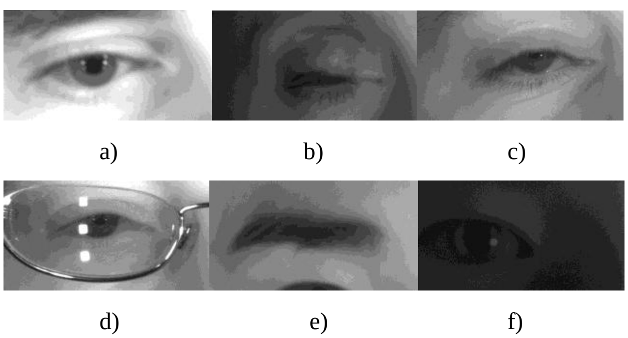


Figure 1. Bad quality eye images: a) Out of focus, b) Closed eye, c) Severely occluded by eyelids, d) Glass and reflection, e) Dark and low contrast.

To the best of our knowledge, this paper is the very first to propose a super-resolution technique for a real low resolution, low quality iris video sequence database - MBGC NIR iris portal database. While all other iris recognition algorithms proposed for MBGC portal database analyse the quality and choose the best quality frame from a portal video sequence for comparison, our approach will fuse information details from all frames to take the advantage of multiple frames in a video sequence.

The remainder of this paper is organized as followed: Super-resolution background is introduced in section 2; section 3 describes the proposed multi-frame robust mean super-resolution for iris image enhancement; section 4 explains our experiments on MBGC database; the paper is concluded in section 5.

2. SUPER-RESOLUTION BACKGROUND

Super-resolution techniques involve three problems: observation model, image registration and reconstruction [6],[7].

Observation model

The first step in the super-resolution reconstruction problems is the formulation of an observation model, which is to develop a model that relates the high-resolution (HR) image to the observed low-resolution (LR) images. Several models have been proposed, but generally the observation model can be expressed as,

$$y_k = DB_k M_k x + n_k \,,$$

where y_k denotes LR images, D is a sub-sampling matrix, B_k is the blur matrix, M_k is the warp matrix, x is the original HR image, n_k is the additive noise that corrupts the image. Various techniques have been proposed for reconstruction HR image from LR images based on above observation model. The key step in the super-resolution process is the registration between the LR source images.

Image registration

Image registration is the process to match two or more images of the same scene taken at different times, from different viewpoints, and/or by different sensors by geometrically aligning the images onto a common reference grid.

Registering images involves defining a mapping or a transformation for pixels from the sensed to the reference image. The transformation can be modeled by the following formula

$$I_2(x, y) = g(I_1(f(x, y))),$$

where $I_1(x,y)$, $I_2(x,y)$ are the pixel values at the relevant locations, f is a transformation that maps the spatial coordinates and g transforms the intensity.

Registration process usually consists of the following three steps:

+ Feature detection and matching: this step identifies salient and distinct features in the reference and sensed images. Matching can be performed through feature-based or area-based. Feature-based methods are suitable for images with bland regions and clearly defined edges or corners, while area-based scheme are useful for highly textured images.

+ Transform parameter estimation: there are two classes of transformation – global and local. Global methods use the same technique for the whole image while local methods treat various regions differently. Global transformations are useful when the scene is relatively static, while local transformations are suitable when objects in the scene move and change independently, such as a surveillance video. There are three main local techniques presented in literature – piecewise interpolation, elastic model-based matching and optical flow.

+ Warping: after the correspondences have been computed, the relative motion between the reference and sensed images can be removed by applying the appropriate mapping to bring the sensed images into reference grid. There are two different warping schemes – backwards and forward warps.

Reconstruction-based super-resolution

Reconstruction-based methods operate directly on the pixel values of the low resolution images without prior knowledge, so these methods are generic. These algorithms can be divided into two classes: frequency domain and spatial domain.

Frequency domain

Frequency domain approaches capitalize on the aliasing that exist in the LR images, an effect easily modeled in the frequency domain. The feature making frequency domain super-resolution algorithms superior to spatial domain methods is their theoretical simplicity. These frequency-based super-resolution methods also have low computational complexity and are suitable for parallel implementation due to the simple decoupling of the frequency domains equations. Moreover, the principal drawback of these methods is their limitation on the type of motion between the LR observations – global translations.

Spatial domain

A numerous amount of spatial domain reconstruction-based methods have been proposed. Examples include non-uniform interpolation, regularized super-resolution reconstruction, projection onto convex sets, hybrid ML/MAP/POCS reconstruction, iterative back-projection and adaptive filtering. These methods try to model a wide range of motions and degradations and to include a prior knowledge for regularization. The flexibility however, comes at the cost of increased computational complexity.

Learning-based super-resolution

While reconstruction-based super-resolution methods try to recover lost high frequency components caused by aliasing, learning-based methods synthesize them instead. A set of training images with high-resolution and corresponding low-resolution image patches is used to provide prior knowledge to reconstruction process. These methods almost always produce visually pleasing images due to the high frequency components created by the process. The problem is that when reconstruction error is high, the resulting super-resolution image is often still a clear image, but it may not look like the original one.

Since learning-based super-resolution techniques may introduce fake high frequencies, our proposal will employ reconstruction-based super-resolution technique to reconstruct a double resolution image from multiple frames of a video sequence.

3. PROPOSED MULTI-FRAME ROBUST MEAN SUPER-RESOLUTION FOR IRIS IMAGE ENHANCEMENT

The super-resolution technique proposed in this paper takes an NIR iris video sequence as input and outputs a super-resolved normalized iris image along with the mask for excluding noise regions. The procedure is illustrated in the Figure 2. The major steps are described as follows:

1. Preprocess the video sequence: Detect and extract the eye region from each frame. Calculate focus level for each frame in the iris video sequence. Exclude unfocused frames. Choose the best focused frame to be the reference frame.
2. Segment iris using two non-concentric circles approximation for pupillary and limbic boundaries for each frame. Then the segmented iris region is normalized using Daugman's doubly dimensionless projected polar coordinate.
3. Interpolate the original normalized iris images and correspondent masks to twice the input resolution using bilinear interpolation.
4. Register the interpolated normalized iris images with the reference image using horizontal translation.
5. Estimate the super-resolved image using robust mean from the reference image and other registered images.
6. Restore the final super-resolved image by applying a deblurring Wiener deconvolution filter.

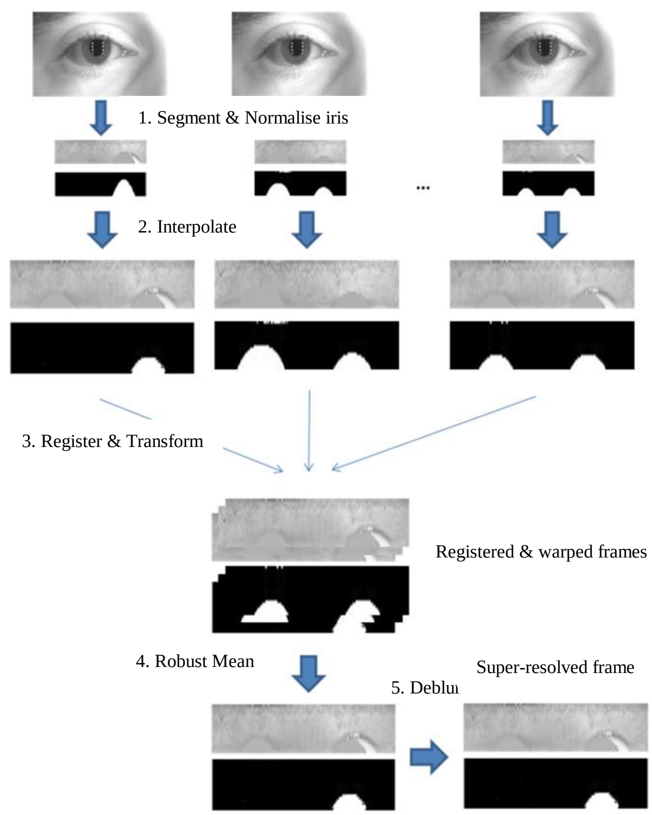


Figure 2. Proposed multi-frame robust mean super-resolution for iris image enhancement diagram.

3.1 Preprocessing

Eye detection and extraction

Eye regions need to be detected and extracted from NIR face video sequence. Here a method called "object detection" proposed by Viola-Jones [8] is employed to identify the eye region in each face NIR frame. Images are represented in "integral" forms which allow the features to be computed very quickly. A learning algorithm based on Adaboost is constructed to select a small number of critical visual features. These learning algorithms are combined in a cascade which allows background regions of the image to be quickly discarded while spending more computation on promising object-like regions.

To detect two-eye region in each frame in NIR face video sequence, we utilise a Haar cascade proposed by Castrillón-Santana *et al.* [9]. The eye-pair classifier has 45x11 pixels size and 19 stages. To improve the searching speed, we define the minimum size for a two-eye region (which is 1000x300 pixels for MBGC NIR face portal database). In addition, to take advantage of continuous video sequence, eye movement between successive frames is estimated to limit the searching region. After the two-eye region is detected from each frame, left eye and right eye are extracted as left and right half.

Focus assessment

Due to short Depth of Field, iris frames captured can be out of focused, which significantly degrades the recognition performance. Severely out-of-focused iris frames need to be eliminated from query images. Defocus primarily attenuates high frequencies components, so focus level can be measured by high frequencies power. Here we exploit the 2-D focus assessment approach proposed by Daugman [10]. A spatial filter is designed to extract the middle and upper frequency band components:

-1	-1	-1	-1	-1	-1	-1	-1
-1	-1	-1	-1	-1	-1	-1	-1
-1	-1	3	3	3	3	-1	-1
-1	-1	3	3	3	3	-1	-1
-1	-1	3	3	3	3	-1	-1
-1	-1	3	3	3	3	-1	-1
-1	-1	-1	-1	-1	-1	-1	-1
-1	-1	-1	-1	-1	-1	-1	-1

Parseval's theorem shows that total power is conserved between spatial and frequency domains

$$\iint |I(x,y)|^2 dxdy = \iint |F(u,v)|^2 dudv .$$

So one frame's total power in frequency domain can be calculated by integrating the power in spatial domain after being filtered by above high pass filter. The focus score can be normalised using the following function

$$f(x) = x^2 /(x^2 + c^2) ,$$

where c is the energy of a clear image. The frames with focus scores which are less than a threshold will be excluded, while the frame with the best focus score in each video sequence will be utilised as the reference frame for super-resolution calculation.

However, that the lighting in a video sequence changes severely from the first frame to the last one causes the focus score unstable. Here, a DCT-based illumination normalisation approach is applied before calculating using Daugman's approach. The final focus score reveals the focus level of the frame. Frames with focus score less than the threshold will be excluded from the query set.

3.2 Segmentation and normalisation

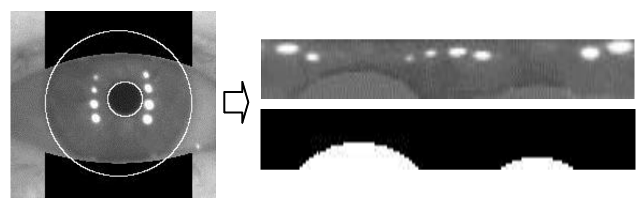
Here iris region is considered to be the region between two non-concentric circles. Inner and outer iris circles are searched using Daugman's approach [10]. Daugman's basic idea of extracting iris from eye image is based on integro-differential operator acting as a circular contour detector

$$\max_{r,x_0,y_0} \left| G_\omega(r) * \frac{\partial}{\partial r} \oint_{r,x_0,y_0} \frac{I(x,y)}{2\pi r} ds \right| .$$

The integro-differential operator search over the whole image domain for the maximum in the blurred partial derivative, with respect to increasing radius, of the normalized contour integral of $I(x,y)$ along with a circular arc ds of radius r and center coordinates (x_0,y_0).

Iris region can be obscured by upper and lower eyelids and eyelashes. The occlusion regions need to be excluded to remain the similarity of the probe image with the gallery image. In our approach, a parabolic curve is fitted into edge images to find upper and lower eyelids.

After segmentation, Daugman's doubly dimensionless projected polar coordinate is exploited to normalise the iris region as shown in Figure 3. This normalisation approach is robust to size-invariant, and pupil-dilation-invariant.



a) b)

Figure 3. a) Iris segmentation using two non-concentric circles approximation, b) Iris normalisation using Daugman's doubly dimensionless projected polar coordinate.

3.3 Robust Mean

Fusing information details from multiple frames effectively is critical to the success of a super-resolution process. In unconstrained environments like MBGC portal database, images can be degraded locally by reflection, eyelids, eyelashes and shadows. These extreme pixels can appear differently in different location in different frames. Patrick Flynn's group proposes fusing by averaging through multiple frames [5]. This approach is only applicable in highly constrained application where all frames are equally good in quality. Even in the highly constrained application, different level of occlusion by eyelids and eyelashes can also corrupt the final fused image. In our approach, robust mean algorithm is used to fuse intensity values of an individual pixel over multiple frames. This algorithm assumes a normal distribution of intensity values, taking the mean of values within two standard deviations from the mean (80% of values). Robust mean fusion scheme is robust against unexpected extreme pixels.

3.4 Deblurring

The final frame after robust mean fusion will be deblurred using a spatially invariant Wiener filter. This spatially invariant Wiener filter reduces the amount of noise present in an image by comparison with an estimation of the desired noiseless signal.

4. EXPERIMENT RESULTS

The experiments on iris verification have been conducted on MBGC portal database. To compare with other methods, we re-implement the Multi-Gallery Multi-Probe method and Averaging method by Patrick Flynn's group [5]. The DET is shown in the Figure 4.

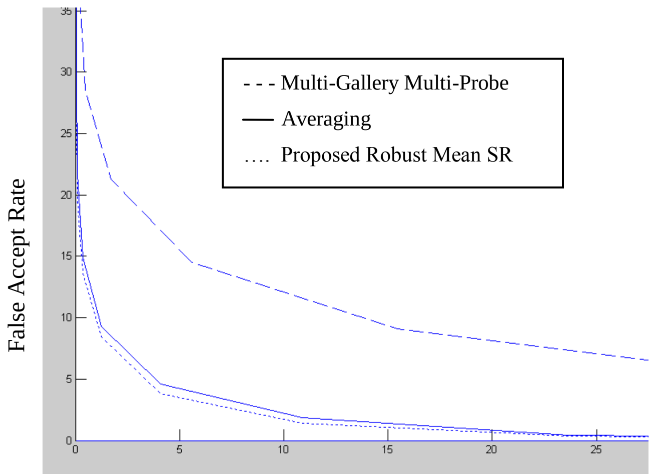


Figure 4. Performance of proposed Robust Mean methods in comparison with Multi-Gallery Multi-Probe and Averaging methods [5].

The DET in Figure 4 shows that the Averaging method [5] performs better than Multi-Gallery Multi-Probe method since a number of outliners have been averaged. Our proposed Robust Mean SR performs better than Averaging method since robust mean fusion scheme is more intelligent than just averaging. The Equal Error Rate (EER) of three methods is shown in the Table 1.

Methods	**EER**
Multi-Gallery Multi-Probe	11.1×10^{-2}
Averaging	4.6×10^{-2}
Proposed Robust Mean SR	4.1×10^{-2}

Table 1. Equal Error Rate (EER) of proposed methods in comparison with Multi-Gallery Multi-Probe and Averaging methods.

In [11], J. Phillips' group has proposed a method for iris verification using best quality frame selection. An EER 4.5×10^{-2} has been reported. Our proposed Robust Mean SR performs better than that.

5. CONCLUSION

In this paper, we have described an effective method for reconstructing a high resolution from multiple frames of a video sequence. Normalised iris images have been used for fusion instead of whole iris images to cope with pupil dilation and eyelid occlusion. A robust mean fusion scheme has been exploited to take the advantage of multiple frames in the video sequence. This pixel-wise fusion scheme performs well on the MBGC portal database in comparison with other methods including Multi-Gallery Multi-Probe and Averaging approaches.

6. ACKNOWLEDGMENTS

Our thanks to University of Notre Dame du Lac, NIST for allowing us to use MBGC database in this research.

7. REFERENCES

[1] A. K. Jain, A. Ross, and S. Prabhakar, "An introduction to biometric recognition", IEEE Trans. On Circuits and Systems for Video Technology, vol. 14, pp. 4–20, 2004.

[2] Kwang Y. Shin, Kang R. Park, Byung J. Kang, Sung J. Park, "Super-Resolution Method Based on Multiple Multi-Layer Perceptrons for Iris Recognition", Ubiquitous Information Technologies & Applications, pp.1-5, 20-22 Dec. 2009.

[3] J. Z. Huang, L. Ma, T. N. Tan and Y. H. Wang, "Learning-based enhancement model of iris", In Proc. of British Machine Vision Conference (BMVC), pp. 153-162, 2003.

[4] Gamal Fahmy, "Super-resolution construction of iris images from a visual low resolution face video", In Proceedings of the 9th International Symposium on Signal Processing and Its Applications, pp. 1-4, 12-15 Feb. 2007.

[5] Hollingsworth; Bowyer; Flynn, "Image averaging for improved iris recognition" Advances in Biometrics. Proceedings Third International Conference, ICB 2009.

[6] Sung Cheol, P., Min Kyu, P., and Moon Gi, K.: 'Super-resolution image reconstruction: a technical overview', IEEE Signal Processing Magazine, 2003, 20, (3), pp. 21-36

[7] Lin, F.: 'Super-Resolution Image Processing with Application to Face Recognition', Queensland University of Technology, 2008

[8] P. Viola, M.J Jones, "Robust real-time face detection", International Journal of Computer Vision, 57(2):151-173, 2004.

[9] Castrillon-Santana, M.1; Deniz-Suarez, O.1; Anton-Canalis, L.1; Lorenzo-Navarro, J.1 Source: VISAPP 2008. Proceedings of the Third International Conference on Computer Vision Theory and Applications, 167-72, 2008

[10] Daugman, J.: 'How iris recognition works', Circuits and Systems for Video Technology, IEEE Transactions on, 2004, 14, (1), pp. 21-30

[11] Yooyoung Lee; Micheals, R.J.; Phillips, P.J.; , "Improvements in Video-based Automated System for Iris Recognition (VASIR)," Motion and Video Computing, 2009. WMVC '09. Workshop on , vol., no., pp.1-8, 8-9 Dec. 2009